Richard Brown CBE
Chairman
Eurostar International Ltd
Times House
Bavingtons Walk
Regents Quarter
London N1 9AW

4 December 2012

Our Ref: Lett6043/Brown/AS/ad

Dear Richard

The Franchising Review and Passengers

When we met in October we promised to follow up with some additional thoughts on franchising and particularly its relationship with, and impact on, passengers.   We have grouped our thoughts under headings of franchise objectives, length and specification but there is inevitably some overlap between them.   We have also developed our thinking on how increasing the size of the National Passenger Survey (NPS) can ensure that passengers have a much bigger role within the franchise process.

Franchise objectives

In July 2010 DfT consulted on reforming rail franchising. It argued that reform had the potential to yield three distinctive types of benefit.

- Better-quality services for passengers
- better value for money for the taxpayer
- create the right conditions for a successful and sustainable rail industry (by providing confidence to invest and giving operators more control over their cost base).

The consultation document pointed out that the interests of these groups need not be mutually exclusive.   We would question, however, whether the right balance has yet been struck.

One of the common criticisms of the franchising process is that bidders construct ever more ambitious bids simply in order to win the franchise. This requires sometimes heroic assumptions on future demand levels which leaves them vulnerable to market fluctuations. The bids for the West Coast franchise show just how high franchise premiums have become. While this looks good from the taxpayer perspective, we worry that it can be false economy from the passenger perspective.   A train operating company (TOC) that is forever playing 'catch-up' on its revenue and demand assumptions may well be driven more by the bottom line than by quality of service.

We would ask whether the bid process takes sufficient notice of these tensions. Does it, for instance, reward bids having high premium/low subsidy more than bids offering greater stability or passenger benefits?

This situation is magnified by the difficulty in forecasting future demand. It is hard to see how any TOC could have accurately predicted events over the past 15 years. The recent recession, for example, was not within the control of the industry nor was it easily predictable. The sheer number of TOCs currently in receipt of 'revenue support' from the Government shows just how difficult this has been – and this in a period of high passenger growth.

While all TOCs face such uncertainty it must be harder for a TOC that has won a franchise with an ambitious bid to weather fluctuations in demand than one that has taken a more cautious approach. There is a school of thought that says that TOCs who take such a risk should bear the consequences. However, this ignores the fact that such risks are also passed on to passengers and staff. Passengers do not want a TOC that is limping along from month to month, raising unregulated fares and car-park prices and cutting future services to make ends meet.

Stability has a value for passengers but it is not clear how this is factored into the current bidding process.

Franchise length
Much of the debate above is wrapped up into the question of franchise length. On the one hand longer franchise terms incentivise the private sector to invest more, thus maximising passenger benefits but on the other they make it harder to forecast demand and quantify risk.

We do not favour a one-size-fits-all approach to franchise length. There will be instances when a single template simply does not fit. We believe that there is a continuing need to look at each franchise on its merits. Among the factors that we feel should influence the length are:

- Plans/desire to merge the franchise with another (having the same end/start date would make this easier)
- The degree of planned disruption envisaged - ensuring continuity of management could help minimise the impact on passengers arising from major upgrade work (e.g. Thameslink).
- Major investment (the argument again being about continuity of delivery)

There may even in some instances be an argument for looking at the performance of the existing operator with a view towards a permanent concession-style of operation, i.e. if performing well, you keep going. Such concessions could support longer-term planning and save cost and management effort in franchise bidding. In theory concessions could even be traded thus creating greater potential for attracting capital investment from the operator. Clearly though, there would need to be some mechanism for incentivising good performance and for mirroring the type of passenger enhancements that arise from the 'taking the franchise to market' model.

However, any move towards longer franchises makes it absolutely essential that there is a robust mechanism for removing poorly-performing operators. It would be unacceptable for passengers to have to wait for the contract to expire before getting rid of an under-performing operator.

One option that may be considered is to split longer franchise periods into 'chunks'. For instance, instead of 15 years you could have two periods – one eight years and one seven (8+7) - or even three successive five year periods (5+5+5). Not only would there be an opportunity to reassess assumptions/models as you approach each new segment but progression could also be made dependent on good performance. If a train company is doing well and meeting various targets then it continues; if it is not then it must either demonstrate how it will improve or risk triggering a new franchise process.

Franchise specification

We believe that the specification is the key to the entire franchising process. We understand the arguments advanced by the rail industry about micro-management hindering the ability of the private sector to innovate. Clearly they would favour as free a hand as possible. However, if nothing is specified how do you monitor a TOC's performance? How, in the worst case, would you remove a TOC for poor performance if there were no standards? How does Government ensure it gets what it pays for with taxpayers' money unless it specifies what it wants in the first place?

So there is a real need for the Government/DfT to continue to have a role in specifying franchises – it should not just be a commercial decision for the operator. Much of the debate, therefore, comes down to the level of specification and the precise targets set within the franchise.

The crucial elements are that the targets reflect passenger priorities and that the measure reflects passenger experiences. We believe that this should take two main forms:

- Hard targets

  There is still a need for the traditional 'hard' performance targets covering punctuality, reliability and crowding. However, we believe that there is a need for much more transparency surrounding these targets.

  Transparency generates greater accountability. Giving rail passengers access to performance figures will help them to hold the train company to account and to ask what is being done to improve services in return for the fares they pay. The 'crowd', in effect, can keep up the pressure on performance on an individual train or route even in the absence of any specific regulatory targets. Good management should not feel threatened by this.

Indeed the availability of accurate data may actually help them – a particularly bad journey can linger in the memory and distort passengers' perceptions.

However, punctuality data is currently only really provided at an overall TOC level which can easily mask significant differences between routes within the same TOC. Providing performance data at a route/service group level would help prevent this and focus attention on areas that need improving. It is important to stress that this information already exists – it would not involve any new costs in terms of data gathering.

Equally, there is currently next to nothing in the public domain about crowding. This is another fundamental aspect of a passenger's journey and an area where greater transparency can again generate improvements for passengers.

The franchise policy review gives an opportunity to address both these issues.

- Passenger Satisfaction

We have long advocated more use of service-quality targets within a franchise. Our strong preference is for targets based on what passengers think – the best judge of quality being those who have used the services in question.

We were pleased that our work on the Southern franchise helped lead to the use of passenger satisfaction targets within the franchise agreement. This established three bespoke targets (for station, train and customer-service attributes) backed up with an enforcement regime that could result in fines being levied if targets were missed.

The National Passenger Survey (NPS) is ideally suited to capture this information. NPS has a large sample size covering some 25,000 passengers each wave. The sampling plan ensures that it is representative of day of travel, journey purpose (commuter, business and leisure), train company and, of course, by a range of demographic attributes (age, sex, ethnicity etc).

We have also worked recently to break down the TOC level analysis into constituent building blocks to provide even more functionality. With a single TOC-wide measure well performing routes can mask poorer performance elsewhere. By breaking overall performance into routes you get a much better indication of where to focus management attention. This greater granularity provides for a subtler, more nuanced franchise measure. However, in order to get maximum value from it there is a need to consider boosting the sample size of each constituent building block. For example, the Cross Country franchise has an overall sample size of around 1000 per wave. We have six constituent building blocks (e.g. Birmingham-Manchester or Birmingham-South Coast) each with a sample size of between 100-200.

The bigger the sample size for each building block the higher the statistical confidence levels. We calculate that increasing sample sizes for each Cross Country building block to around 800 would effectively double the accuracy/half the confidence interval. We attach a more detailed technical analysis looking at sample sizes and confidence intervals. We also look at the associated costs of doing so across all operators: we estimate that it

would cost in the region of £500,000 (including VAT) if franchise targets were to be based on a single wave of NPS and £250,000 if two waves were to be combined as now.

We believe a franchise should have such targets and that they should count towards any assessment of a TOC's suitability to continue (as per the section above). NPS data can also be used to weight the importance of franchise targets so that they represent what is key for passengers. Not only does this have the added benefit of making passengers part of the decision-making matrix (where previously they had no such role) but it also promotes a greater sense of accountability between service provider and consumer.

Moreover, the use of such targets can be framed in a way that is entirely consistent with Government policy and which still gives the train company considerable leeway to act. For instance, rather than specifying that a franchisee purchase 50 new ticket-vending machines (an input target) it could require it to increase passenger satisfaction with retailing (an output target).

There is also the potential to use passenger satisfaction as an incentive mechanism which rewards TOCs that exceed targets rather than just as part of a penalty regime.

We would be happy to discuss these issues in more detail.

Yours sincerely

Anthony Smith
Chief Executive

NPS sample sizes and confidence intervals – technical summary

NPS targets are being increasingly incorporated into franchise agreements with train companies (TOCs), and generally these targets are based on routes or groups of stations (called building blocks). There are 19 franchised TOCs and these are split into 74 building blocks.

For individual building blocks and TOCs the 95% confidence interval for the satisfaction scores will vary depending on the sample size and percentage of passengers satisfied for any particular factor.

The tables below show how the 95% confidence intervals vary for two specific TOCs (CrossCountry and Greater Anglia) given the sample sizes achieved for an individual wave (spring 2012) and three different figures for the percentage of passengers satisfied (50%, 70% and 90%). If passenger satisfaction is 50% this is the maximum width of the 95% confidence interval.

TOC/Building Block	% Sat			Current sample sizes
	50%	70%	90%	
CrossCountry TOC Overall	3.2%	3.0%	1.9%	1000
Cross Country - Birmingham – Manchester	9.0%	8.2%	5.4%	100
Cross Country - Birmingham - North East & Scotland	6.6%	6.0%	3.9%	250
Cross Country - Birmingham - South Coast	7.5%	6.9%	4.5%	200
Cross Country - Birmingham - South West	8.3%	7.6%	5.0%	175
Cross Country - Birmingham – Stansted	7.8%	7.1%	4.7%	150
Cross Country - Nottingham – Cardiff	9.5%	8.7%	5.7%	125
Greater Anglia TOC Overall	2.4%	2.2%	1.4%	2000
Greater Anglia – Intercity	5.8%	5.3%	3.5%	400
Greater Anglia – Mainline	5.1%	4.7%	3.1%	400
Greater Anglia – Metro	5.2%	4.8%	3.1%	400
Greater Anglia – Rural	8.5%	7.7%	5.1%	200
Greater Anglia - Stansted Express	7.9%	7.3%	4.8%	200
Greater Anglia - West Anglia	5.2%	4.8%	3.1%	400

This table shows that for example assuming satisfaction for CrossCountry of 70% the 95% confidence interval is ±3.0%. 70% is close to the average for the percentage of passengers satisfied taking all station and train factors together (though there is quite wide variations both by factor and by TOC).

Note: Confidence intervals assume that all (or nearly all) passengers answer a particular NPS question. For a few questions, for example, 'how well train company deals with delays' this is not the case and therefore the confidence interval will be wider.

To double the accuracy (or halve the confidence intervals) of the estimates, the sample size needs to be increased by a factor of 4 (e.g. from 200 to 800).

Building-block confidence intervals of ±4% per wave for individual factors
Assuming 70% of passengers are satisfied, to achieve a 95% confidence interval of ±4% per wave the sample size for each building block (based on the sample sizes achieved in spring 2012) needs to be increased by approximately 2,750 more respondents for CrossCountry and 2,100 for Greater Anglia.

Nationally the sample size for all franchised TOCs would need to be increased by approximately 28,109 respondents (i.e. more than doubled) for each building block to have a ±4% confidence interval (assuming 70% satisfaction) each wave. The cost of this increase in sample size would be, based on contracted 2013 prices per returned questionnaire, approximately £500,000.

Building-block confidence intervals of ±4% based on latest two survey waves for individual factors
The target scores in franchise agreements however are based on two waves combined. This has the dual advantage of both doubling the sample size (to increase accuracy) and including both a spring and an autumn wave, and thus balancing out the seasonal effect we see across the year.

To achieve a 95% confidence interval of ±4% for individual factors for two waves combined for each building block, the sample size would need to be increased by approximately 13,654 nationally (cost c. £250,000).

Building-block confidence intervals of ±2% based on latest two survey waves and combined factors
Current and proposed targets in franchise agreements also combine several NPS train and station together into composite measures (for example Station factors, Train Service factors, Train Comfort factors and Customer Service factors).

By combining several station and train factors together in the calculation of the target-scores, the confidence intervals are much narrower. Calculation of target score confidence intervals is much more complex to do and is partly dependent on which factors are proposed for inclusion, but based on the methodology proposed for the Essex Thameside, Greater Western and Thameslink franchises the confidence intervals for each of the routes have been calculated.

Due to the nature of the individual franchises and historic sample sizes, Essex Thameside and Greater Western confidence intervals for each of the building blocks for Station, Train Service, Train Comfort and Customer Service are all currently within ±2%. On Thameslink, however, confidence interval targets for some building blocks for 'Train Service' and for all building blocks for 'Customer Service' exceed the ±2%. Confidence intervals for most building blocks for most TOCs would also currently exceed ±2%.

In order to achieve the ±2% threshold then an increase in the sample size for Thameslink will be required given the current NPS target specifications. The sample size increase needed is approximately 6,000 over two waves. This is therefore an increased sample size of approximately 3,000 per wave. Most of the increased sample size is on the Thameslink South building block.

Different percentage thresholds can be calculated if a percentage other than ±2% is seen as appropriate.

Assuming the same list of factors that were included in the methodology proposed for the Thameslink franchise, the sample size increase required for all building blocks for all TOCs to not exceed ±2% is approximately 44,600 annually or 22,300 per wave. The cost of such a sample size increase (at 2013 prices) is approximately £400,000.

The figures quoted are 'ball parks' as they contain some assumptions about average performance across TOCs in order to reduce the number of calculations necessary. If there was interest in taking these ideas forward we could work up more detailed numbers.